Learning of User Formulations for Business Listings in Automatic Directory Assistance

(Article begins on next page)

Learning of User Formulations for Business Listings in Automatic Directory Assistance

C. Popovici, M. Andorno, P. Laface Γ, *L. Fissore, M. Nigra, C. Vair* Κ

Γ Politecnico di Torino, Italy
{andorno,laface}@polito.it
Κ Loquendo, Torino, Italy
(Cosmin.Popovici,Luciano.Fissore,Mario.Nigra,Claudio.Vair}@loquendo.com

Abstract

Automatic Directory Assistance (DA) for business listings poses many application specific problems. One of the main problem is that customers formulate their requests for the same listing with a great variability.

This paper presents the results of a study aiming at the evaluation of an approach towards automatic learning, from field data, of expressions typically used by customers to formulate their requests for the most frequent business listings.

We use a clustering procedure that exploits the association of the phonetic string produced by a lexical unconstrained search for a given denomination pronounced by the user and the phone number provided by the system or by the human operator, in case of failure of the automatic DA service.

We show that an unsupervised approach allows to detect user formulations that were not foreseen by the designers, and that can be added, as variants, to the denominations already included in the system to reduce its failures.

1. Introduction

Telecom Italia, has deployed an automatic DA system. The automated DA service, developed by Telecom Italia and Loquendo (formerly CSELT), routinely serves customers asking for both residential and business listings. Whenever the automatic system is unable to terminate the transaction with the customer, the call is routed to a human operator.

The Italian telephone book listings includes more than 25.000.000 records, about 3.500.000 of which are business listings. Since about 80% of the DA accesses is related to business listings, it is important to improve the percentage of success of the automatic DA system for this class of calls. The design of an automatic DA system for business listings poses several problems, one of the hardest to be solved is that customers formulate their requests for the same business listing with great variability. Several approaches have been considered to face this problem. By using statistical language modelling and continuous speech recognition technology a quite flexible system with respect to the user formulations, providing high linguistic coverage, could be developed. However, since the perplexity of the task is very large, the risk of obtaining poor performance with a weakly constrained speech recogniser is relevant. Furthermore, it is more difficult to develop reliable confidence measures and rejection strategies for continuous speech. A word spotting approach presents the same problems enhanced by the size of the vocabulary.

It was decided, thus, to design the DA service using a large vocabulary isolated word recognition technology, where the sequence of words in business listings is concatenated and transcribed as a single word, with possible silences in between. Since the content of the original records in the database does not, typically, match the linguistic expressions used by the callers, a complex processing step is needed for deriving a set of possible formulations variants (FVs) from each original record in the book listings.

The advantage of the FV approach is to supply the speech recognizer with some knowledge about the variability of the user formulations, and to allow the use of isolated word recognition technology with the capability of dealing with out of vocabulary words. On the other hand, it is clear that a large percentage of expressions will not be perfectly covered by the FV database, and that the complexity of the search increases because the size of the system vocabularies increases.

This paper presents the results of a study aiming at the evaluation of an approach towards automatic learning, from field data, of expressions typically used by customers to formulate their requests for the most frequent business listings.

We partition the field data referring to a given denomination into phonetically similar clusters from which new user formulations can be derived. Our clustering procedure exploits the association of the phonetic string produced by a lexical unconstrained search for a denomination pronounced by the user and the phone number provided by the system or by the human operator in case of failure of the automatic DA service.

The paper is organized as follows: Section 2 gives a short overview of the Loquendo DA system. Section 3 recalls the steps for obtaining FVs from the records in the book listings and from field data. Section 4 details our approach for learning new formulations from field data, and discusses the obtained results. Finally, our conclusions are given in Section 5.

2. Loquendo DA system overview

As introduced in the previous section, large vocabulary isolated word recognition is the basic technology embedded in the Loquendo DA application.

This work was partially supported by the EU ITS Project SMADA Speech Driven Multi-modal Automatic Directory Assistance

Pref: 011 *Tel*: 5175296
City: TORINO *Prov*: TORINO
Address: 63, C. VITTORIO EMANUELE II
***Den*: BAR RISTORANTE LA FORCHETTA D'ORO DI MARIO ROSSI**
***Description*: PIZZA**
***Category*: RISTORANTI**

RISTORANTE LA FORCHETTA D'ORO
BAR LA FORCHETTA D'ORO
BAR RISTORANTE LA FORCHETTA D'ORO
PIZZERIA LA FORCHETTA D'ORO
LA FORCHETTA D'ORO

Figure 1*:* Example of some fields in a record, and its formulation variants

The first recognition step decodes the user utterance by means of a Hybrid HMM-NN model, where the emission probabilities of the HMM states are estimated by a Multi Layer Perceptron. The best unconstrained phonetic string and its score are also generated in this step.
The second step, based on Continuous Density HMMs, decodes the same utterance using as a vocabulary the N-best hypotheses produced by the first step. The added value of this second step is twofold since the combination of the hypothesis scores of the two steps, not only increases the recognition accuracy, but allows also the production of a reliability score for the best hypothesis. This score, together with the phonetic one, is used by the dialog manager module for rejecting unreliable hypotheses or for reducing unnecessary request turns.

3. Generation of formulation variants

The description of the steps for generating the formulation variants from the original book listing records is out of the scope of this paper. It will suffice to say that the final result of these steps is a semantic table, obtained by applying rules specific of each business category, that summarizes the content of the record, reducing the variability of the information inserted by the operators in the record fields. Using the semantic table information, it is possible to generate a set of formulation variants with an associated score. An example of record, and its generated formulation variants, ordered by score, is shown in Fig. 1. The best scoring formulation is also played back to the customer for confirmation.
Several turns for evaluating the coverage of the user formulations by the FVs were performed. In a first phase real user data were collected from the interactions with human DA operators located in Turin, then from calls to a prototype DA serving the Catania telephone district. From these preliminary tests it has been verified that the coverage of the original FVs was about 40%. It was, thus, mandatory to generate more accurate formulations for frequently requested listings, in particular for those presenting high failure rates.
Another large database (DB20000) was then collected from a month and a half of customers calls to an automatic system operating in Rome during the night. In particular, 8848 business calls, routed to the human operator by the system, because it was unable to deliver the desired information, were selected and transcribed. Another set was selected, and transcribed, from the daily traffic managed by 13 call centers distributed in several regions of Italy. All these calls correspond to the most frequently asked listings. The database includes a total of 20216 transcribed calls associated to the phone number provided by the human operators.
To generate new, more accurate, FVs, the transcribed denominations were analyzed, and generation rules derived, depending on the business category, according to a priori knowledge and data evidence. The FVs that received most attention were those related to hospitals, social services, public utilities, communication and transportation agencies, and the like, because they account for the majority of the calls. Since the automatic DA system is currently fully operational, new FVs, and possibly rules, are also derived whenever the service provider signals consistent and important anomalies.
By using the FVs rules derived from this new field data, the coverage of the FVs increased from 40% to more than 60%, using an average of 5 FVs per denomination. This also means that many users are rather collaborative and that the system prompts elicit concise linguistic expressions.

4. Automatic learning of formulation variants

An analysis of the DA system failures has been done to discover the main causes of its errors. The errors can be grouped in three main classes:

1. User formulations are slightly different (due to articles, prepositions, etc.) with respect to the stored set of FVs.
2. User formulations are different with respect to the stored set due to the insertion of extra words or sentences or due to the deletion of words, even though part of the information is still there.
3. User formulations are completely different with respect to the stored set.

We focus, in this work, on the errors of the first and third class. In particular, frequent errors of the third class for a specific entry can give an indication that the insertion of new formulations in the DA database is required for that entry.

4.1. Phonetic transcription

From the calls routed to the operators, the list of the most frequently requested phone numbers (provided by the operator) was selected. To each phone number the trace and the recording of the corresponding call has been associated.
Setting a threshold of 20 requests per phone number, the most requested listings for the 3434 calls in the Catania database are 16 only. A much higher spreading has been experienced, as expected, for the nationwide calls, where 53 listings only were requested more than 20 times.
As said in Section 2, the recognition module of the system produces, together with the lexical constrained word hypotheses, the phonetic transcription of each utterance as the best sequence of phones obtained using a looped phone model.
The phonetic strings associated to a given phone number are, thus, the automatic transcriptions of the different way in which users formulate their request for the corresponding business listing.
Table 1 shows, as an example, a small set of unconstrained phonetic transcriptions associated to the most requested phone number in the DB20000 database: *848888088* corresponding to *FSInforma*, a widely used automatic train timetable information system, developed by CSELT, and managed by the Italian railways service provider *Ferrovie dello Stato.*
These phonetic strings are widely different, and some of them can hardly be decoded. Recall, please, that these utterances were not completed by the automatic DA system for several

ufiCoinfoRmaZionilstaZiuneditaeni	
enomaladelataZaneditoRenopoRtanovelomRaveRda	
feRoviodelostato	
ifuoRmaZionifeRoiedelostato	
esaZionetiboRtina	fveRuilstato
skaZione	oRaRiodetReni
tReno	fRovionalostato
nomaRomeRbefese	saZinoCentRale

Table 1: Samples of transcriptions of user requests for the railway information service *FSInforma*

reasons such as endpoint detection failures, extra-linguistic phenomena, low confidence scores, recognition errors due to the lack of a suitable transcription in the current database, etc. Another cause of system failures is that the user request was ambiguous, incomplete or embedded in a sentence, so that only several turns of dialog with the user allowed the operator to deliver the information.

On the other hand, in Table 1 it is possible to detect phonetic sequences that are easily interpreted since they are correct or nearly correct transcriptions of a denomination such as <feRoviodelostato> and <skaZione> for “Ferrovie Dello Stato” and “Stazione” respectively, and several variants with relatively few phonetic distortions.

It is also worth noting that, given a huge number of requests for the same phone number, there is a high probability of obtaining clusters of phonetically similar strings. The measure of the distance between two strings of phones can be obtained by Viterbi alignment of the two strings using the inverse of the log-probability of insertion, deletion and confusion among phones. These probabilities were trained using another set of field data, aligning each phonetic sequence with its corresponding correct transcription. These data were also exploited for training field adapted acoustic models.

4.2. Clustering and selection of new formulations

For the most frequently requested phone numbers, each set of phonetic strings was clustered into similar subsets by using a furthest neighbor hierarchical cluster algorithm based on the mutual distance between each phonetic string.

The set of phonetically similar utterances is detected by stopping the bottom-up clustering when the number of elements that have been partitioned is greater than $N-N/(\log_2 N + 1)$, where N is the number of strings that are clustered. The clusters with few elements and large within cluster variance are discarded.

An interesting example of output of the clustering procedure is related to the main Catania Hospital, whose phone number is requested using several formulations referring to different clinics within the same hospital. Six denominations were added manually as formulation variants after the preliminary analysis of the errors of the prototype system in operation in Catania, The clustering algorithm detected the same formulations: a few elements of two clusters are shown in Table 2.

Another example of automatic clustering, detecting a pronunciation variant of a foreign denomination, is reported in Table 3, where some phonetic strings of two clusters related to requests for the French word *Auchan* are shown. As can be argued, the number of available samples for the Catania database is too small for deriving reliable phonetic transcriptions for new formulations. However, if a large enough database is available, it is possible to select significant clusters,

Ospedale Santa Marta	Ospedale Ferrarotto
<t&>ospedalafantamaRta	topelaleseRaRato
<t&>ospedalafantamaRta	sospelalefeRaRatoi
fospedalesantamasta	ostedavefeRanoto
	osbedaRefeRalato

Table 2: Two clusters for different formulations related to main Catania *Hospital*

French pronunciation with final schwa	Italian pronunciation
o&ana	au&an
vo&ana	au&an
fo&ana	auv&a
fo&an	aui&ian
o&a	au&aen

Table 3: Samples of two clusters of similar pronunciations for the denomination *Auchan*

Central element	Within - system nearest variant	No of elements	Cluster variance	Distance
feRoviedelostato	feRoviedelostato	156	2.13	0.00
staZioneCentRale	staZionefeRoviaRia	198	3.27	3.22
staZione	staZionefeRoviaRia	25	1.9	4.43

Table 4 – Central elements of the three significant clusters related to the denomination *FSInforma*

characterized by high cardinality and small dispersion of the included phonetic strings. For example, using the 458 formulations that were available for the phone number of the *FSInforma* in the DB20000 database, the procedure generated several phonetically similar clusters, but only three of them were significant according to a selection criterion related to the number of elements in the cluster (> 20 in this case) and to a low (< 4.0) dispersion of the elements within the cluster. The central element of the three clusters, defined as the string that has the minimum sum of the distance from all the other elements of the cluster, is shown in Table 4.

It is worth noting that, when the number of elements collapsed into a cluster is large enough, the central element of the cluster gives a very good transcription of the required denomination.

For the central elements in Table 4, good formulation variant candidates are the phonetic strings <staZioneCentRale> and <staZione> that are quite distant from the already present formulation <staZionefeRoviaRia>, while <feRoviedelostato> exactly matches a formulation already in the system.

4.3. Assessment of the procedure

As stated in Section 4.1, the Catania and DB20000 databases were processed to include the phone numbers provided by the human operators. To assess the capabilities of our approach we need a huge amount of phonetic strings and their associated phone numbers. Since it is, currently, cumbersome to associate the phonetic strings with the phone number that will be eventually delivered by the operator, rather than relying on the calls that the automatic DA failed to serve, we cluster the phonetic strings of calls successfully processed by the automatic system. Using these data, available at our will, we can assess the quality of our clustering procedure, and its ability to produce, as a central element of a cluster, a formulation variant that has been included in the system.

During February 2001 it was very easy to collect, from a platform located in Turin, serving part of the North-West region

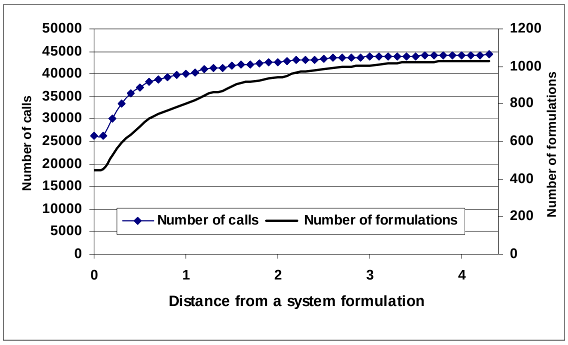

Figure 2: Number of formulations (calls) matching a system formulation variant within a phonetic distance

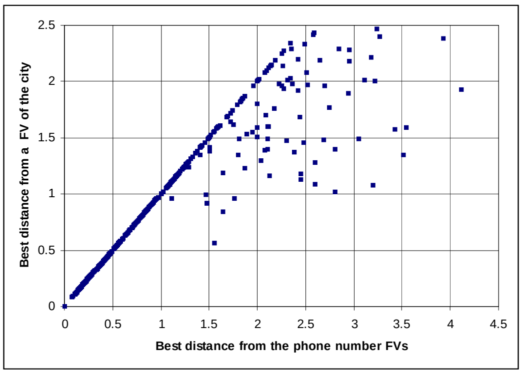

Figure 3: Scatter plot of best distances of a central element of a cluster from the FVs of its phone number (x-axis), and from all the FVs of the city (y-axis)

of Italy, 336113 phonetic strings related to business listing calls, and the related phone number provided by the automatic DA service. There are a total of 153043 different listing references in this database (DBFEB01), 108026 of which have been requested only once. To obtain reliable statistics we processed the 358 denominations that were requested more than 50 times, for a total of 46516 calls. Again, the most frequently requested one is the *FSInforma* service, and the majority of the other calls refer to hospitals, police, cinema, public offices, television, and transportation.
Figure 2 shows the number of formulations (calls) that match a system FV, corresponding to that phone number, within a phonetic distance. It can be observed that 448 automatically detected FVs (43.8%) perfectly match a FV that was included in the system. These formulations cover 26200 (56.3%) requests for 270 different denominations (75.4%). 803 formulations (77.9%), covering 40079 calls (86.1%) for 335 denominations (93.6%), are within a distance of 1.0 from the nearest system FVs. This distance corresponds to slight phonetic variations as the ones shown in the first row of Table 5.

Cluster central element	Cluster variance	Distance	System formulation variant
kameRadikomeRCo	0.73	0.94	kameRekomeRCo
ospedalebesta	0.83	2.28	ospedale
atenaseRviZi	1.09	2.06	atenaesepia

Table 5: Examples of detection of new formulations

These results demonstrate that the cluster algorithm is able to detect very well most of the old formulations. In particular, the example in the first row of the Table 5 shows that users correctly pronounce <kameRadikomeRCo>, including the Italian preposition "di" (of) that does nor appear in the current FVs.
Moreover, the automatic DA service may succeed in completing a user request, even though his formulation is distant from a system FV, because the dialog manager may ask the address information to the user. Thus, the DBFEB01 database includes associations of the correct phone number with phonetic strings that are largely distant from the nearest system FV, but are correct transcriptions of new formulations. Example of these formulations are shown in the last two rows of Table 5: in the third row the name of the hospital "*Besta*" is normally formulated by the users, while the acronym S.P.A. is substituted with word "*Servizi*" (services).

As a final assessment we computed, for each of the 1031 central elements produced by our clustering procedure, its distance from *all* the system FVs of the city corresponding to the area code of the phone number associated with the cluster. Figure 4 is a scatter plot where the *x* and *y* coordinates of a point represent the distance of a phonetic string from the nearest phone number FV, and from the nearest FVs of the city listings respectively. Thus, the 934 points laying on the 45° right line correspond to automatic phonetic transcriptions that are close to one of the FVs associated to the correct denomination. As better shown in Fig. 2, 448 of them perfectly match a system FV. The 97 points laying below the 45° right line indicate that a central element is closer to an incorrect denomination. It is worth noting, however, that the points located on the upper right side of the figure may correspond to new formulations if the cardinality of the corresponding cluster is high and the within cluster variance is low. We obtained, for example, the phonetic string <CentRokomeRCalemetRopoli> as the central element of a cluster including 48 user formulations, having a cluster variance of 0.94, and a distance 3.54 from the current best system FV <CentRokomeRCale>. Since it satisfies the criterion of cluster robustness and "purity", and since it is far from the nearest system FV, it can be added to the current FVs as a new formulation.
The distribution of the points in Fig. 3, concentrated on the (lower part of) the 45° right line, assesses the viability of the approach for incremental learning of new (phonetically accurate) formulations if a large enough database is available for the requests routed to the operators.

5. Conclusions

We have shown that an unsupervised approach is able to detect user formulations that were not foreseen by the designers of a DA system. These formulations can be added to the system to reduce its failures. Conversely, the system formulations, generated by the rule-based system from the book listings for a given denomination, that never appeared in a huge amount of real data can be purged.

6. References

[1] R. Billi, F. Canavesio, C. Rullent, Automation of Telecom Italia Directory Assistance Service: Field Trials results, Proc. IVTTA 1998, Turin, pp. 11-16, 1998.